

Илья Иванченко
Магнит



Agenda

- Как подружить нейросети с временными рядами
- Как мы выстроили пайплайн
- Ожидаемый и реальный эффект
- Анализ скрытых представлений



О бизнесе



6

форматов

>25к

магазинов

2–25к

SKU в ассортименте

>150к

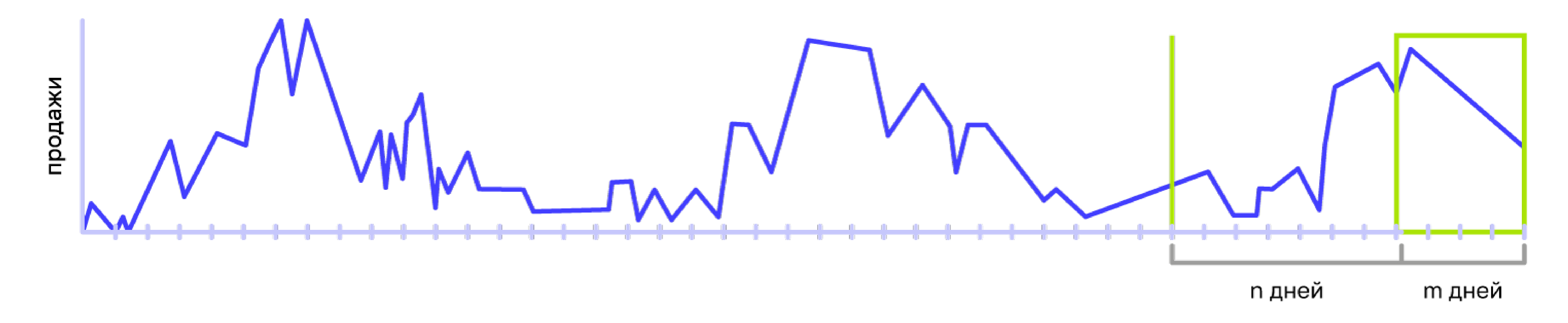
SKU всего

*по данным на ноябрь 2022 г.



О задаче

- Формирование прогноза спроса для каждой комбинации магазин-товар-день
- Горизонт прогноза нефиксированный: **от 1 дня до 1 года**
- Плечо прогнозирования является нефиксированным и составляет **~ 30–180 дней**





Предпосылки

- Классический стек уже реализован
- Отсутствие возможности полноценно учитывать происходящее до и после целевого периода
- Наличие большого количества моделей в текущих подходах

Базовые подходы

Кастомные алгоритмы

Бустинги

Нейросети

Данные – наше все

Неважно, что обучать –
важно, чем кормить

Вклад от развития модели

Вклад от проработки данных



Типичные данные

Аномалии

Виртуальные остатки

Пропуски во временных рядах

Искаженный спрос

Обнуления

Искаженный спрос

Виртуальный остаток

Пропуски

Временное закрытие

Промо

Аномалии



Объемы

1 Тб
в модель

200 Тб
Data Lake

700 Тб
DWH

Более **50 млрд** наблюдений за 1 год



NN и time series

Отсутствие популярности у нейросетей для табличных данных

Малое количество общедоступной экспертизы

Google

the most popular neural networks

Все · Картинки · Видео · Новости · Покупки · Ещё · Инструменты

Результатов: примерно 348 000 000 (0,46 сек.)

Top 10 Neural Network Architectures in 2022 ML Engineers Need to Learn

- LeNet5.
- Dan Ciresan Net.
- AlexNet.
- Overfeat.
- VGG.
- Network-in-network.
- GoogLeNet and Inception.
- Bottleneck Layer.

Ещё

https://www.upgrad.com › Blog › Artificial Intelligence

Top 10 Neural Network Architectures in 2022 ML Engineers ...



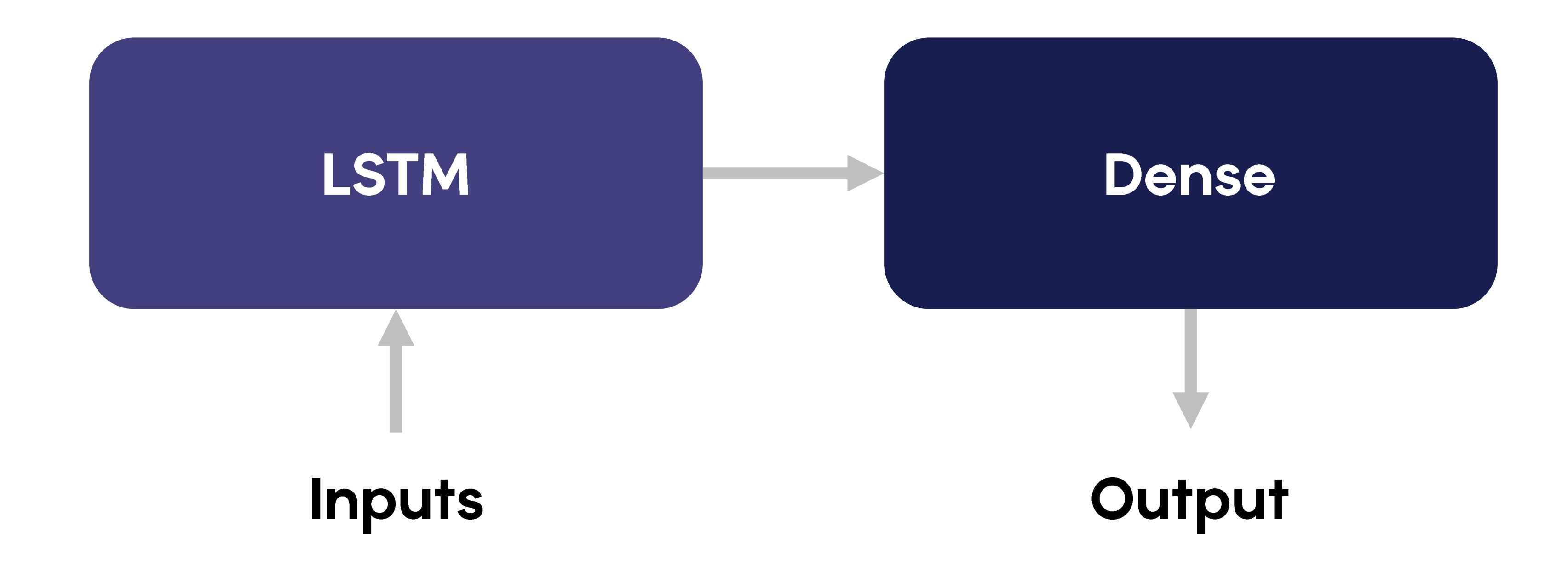
Сложности на пути

- Дороговизна
- Большая требовательность к подготовке данных
- Экстраполяция
- Черный ящик для бизнеса
- Более низкие метрики



Базовые элементы

Для работы с последовательностями

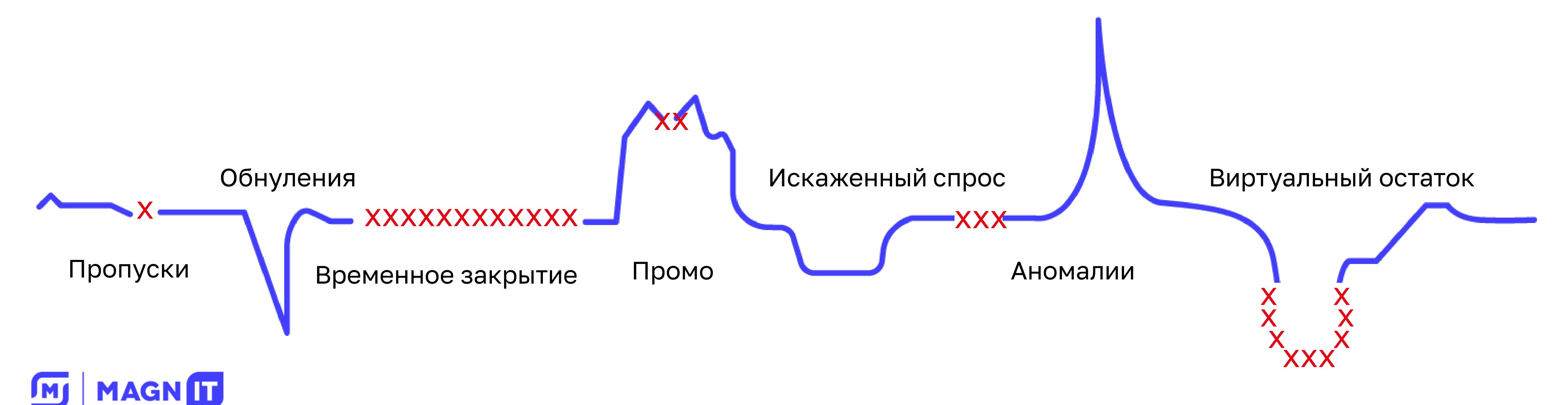


Маскирование

Преимущества:

Цельный временной ряд

Отсутствие фиктивных значений



Маскирование

Преимущества:

Цельный временной ряд

Отсутствие фиктивных значений





Неизвестные факторы

Признаки

Наблюдения

День	Скидка	Продажи
k	0%	-
...	...	
0	**Дата расчета**	
-1	0%	4
-2	0%	2
-3	0%	3
...	...	
-n	10%	7

avg

В случае работы
в скалярном формате
одно наблюдение –
это одна строка



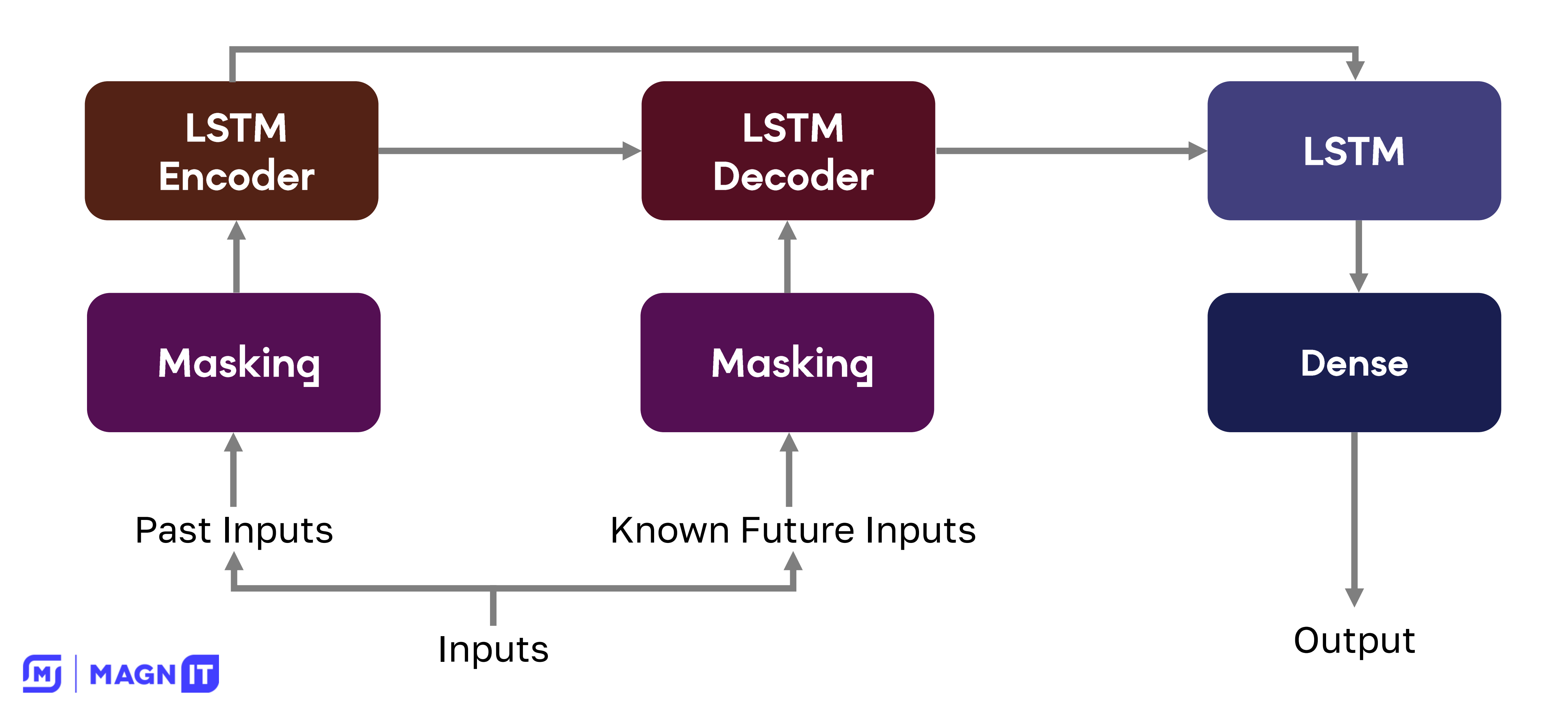
Неизвестные факторы

	День	Скидка	Температура	Продажи
Будущее	k	0%	-	-
	...	...		
	3	20%	-	-
	2	0%	-	-
	1	10%	-	-
	0	**Дата расчета**		
Прошлое	-1	0%	10	4
	-2	0%	12	2
	-3	0%	13	3
	...	...		
	-n	10%	10	7

В случае работы **с последовательностями** одно наблюдение – это одна таблица



Неизвестные факторы



Статичные признаки

	День	Торговая площадь	Скидка
Будущее	k	364	0%
	...	...	
	3	364	20%
	2	364	0%
	1	364	10%
	0	**Дата расчета**	
Прошлое	-1	364	0%
	-2	364	0%
	-3	364	0%
	...	...	
	-n	364	10%

Преимущества:

- Корректное использование статичных признаков
- Оптимизация используемой памяти

Статичные признаки

LSTM Encoder
LSTM Decoder
LSTM
Masking
Masking
Dense
Static data
Past Inputs
Known Future Inputs
Inputs
Output



Embeddings

Преимущества:

- Минимизация фиче-инжиниринга
- Работа с высокоразмерными кат. фичами

Input Encoding

0
1
0
0
...
0

Dense vector

0.3
0.1
0.9
...
0.0

Categorical features → Input Encoding → Embedding model → Dense vector → Model (Deep learning)

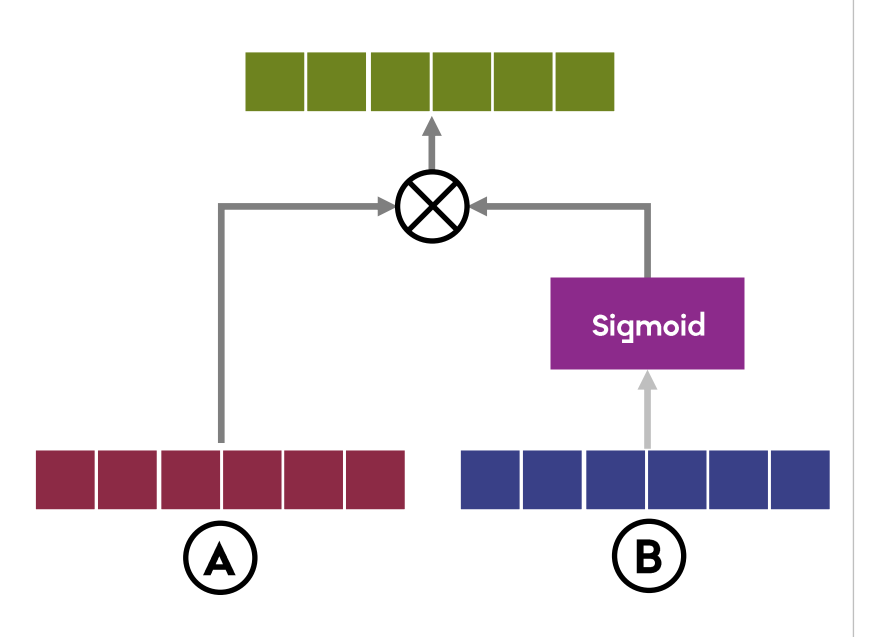


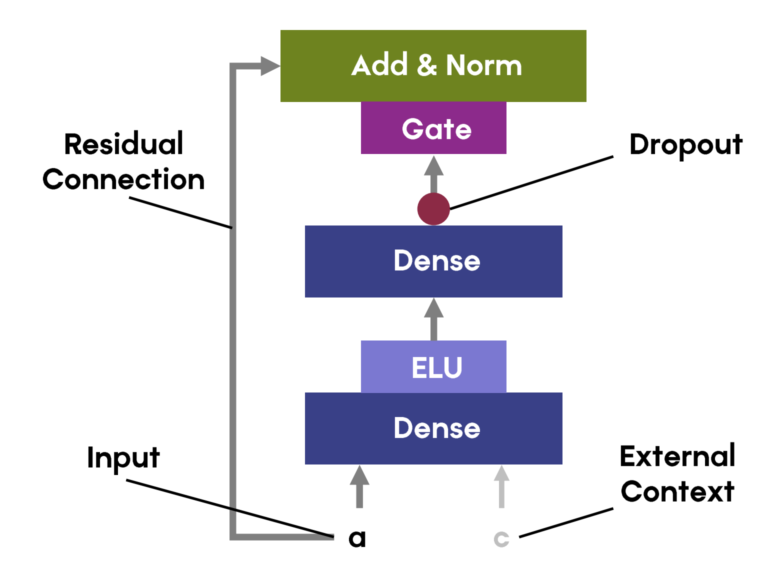
Категориальные факторы

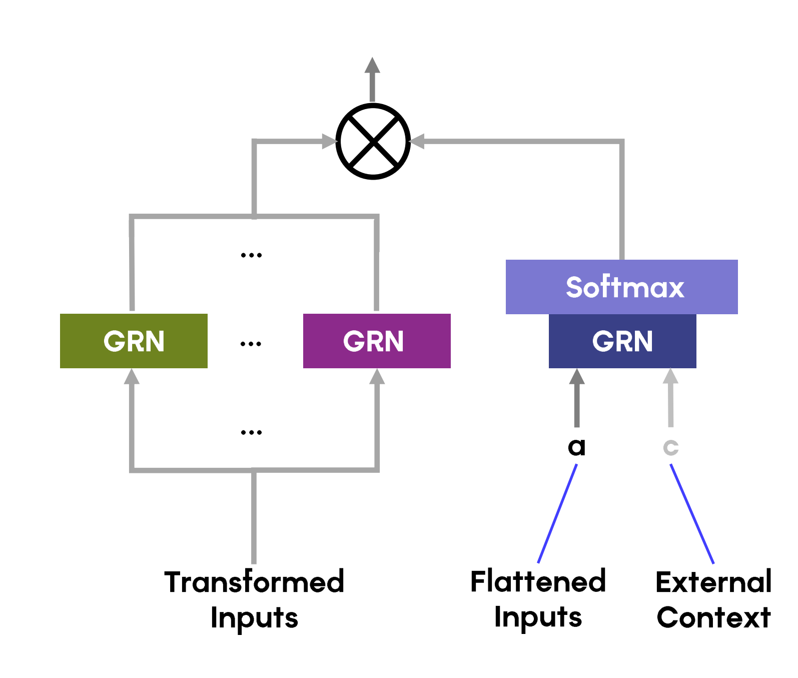
LSTM Encoder
LSTM Decoder
LSTM
Embedding model
Embedding model
Dense
Masking
Masking
Static data
Past Inputs
Known Future Inputs
Inputs
Output



Variable Selection Network







Gated Linear Unit (GLU)

GLU помогает в выборе наиболее важных признаков

Gated Residual Network (GRN)

Помогает сети понять, какие входные преобразования просты, а какие требуют более сложного моделирования

Variable Selection Network (VSN)

Представляет собой механизм выбора признаков и позволяет различать полезные факторы от зашумленных, используя для этого GLU и GRN

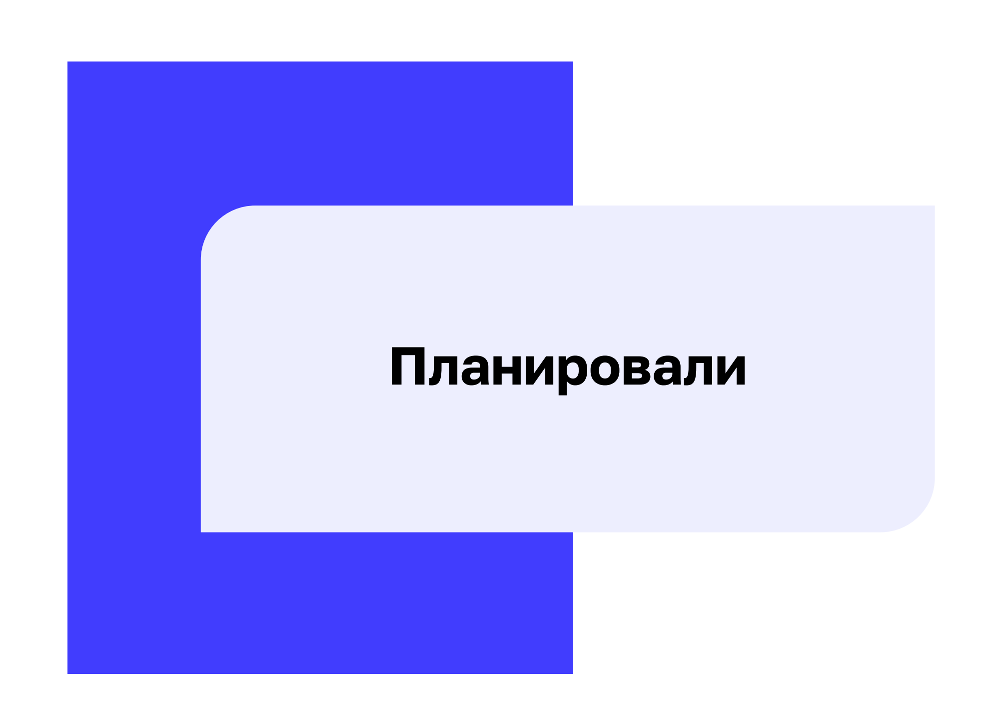


Модель

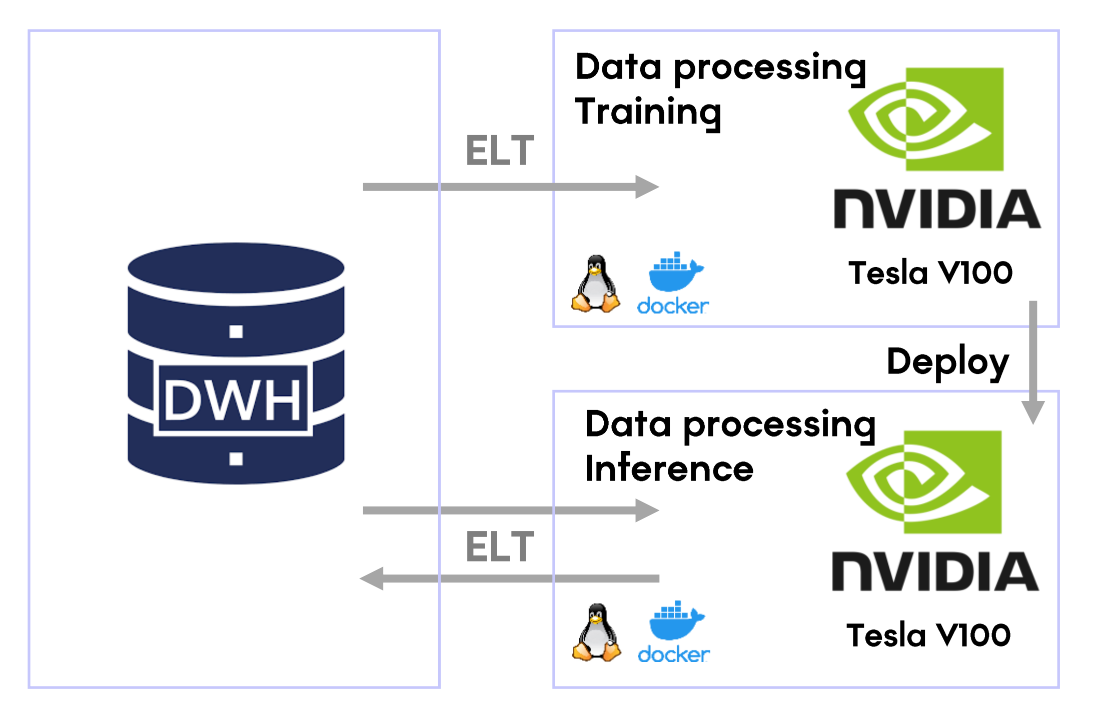
LSTM Encoder
LSTM Decoder
LSTM
Variable Selection
Variable Selection
Variable Selection
Embedding model
Embedding model
Dense
Masking
Masking
Static data
Past Inputs
Known Future Inputs
Inputs
Output



Пайплайн

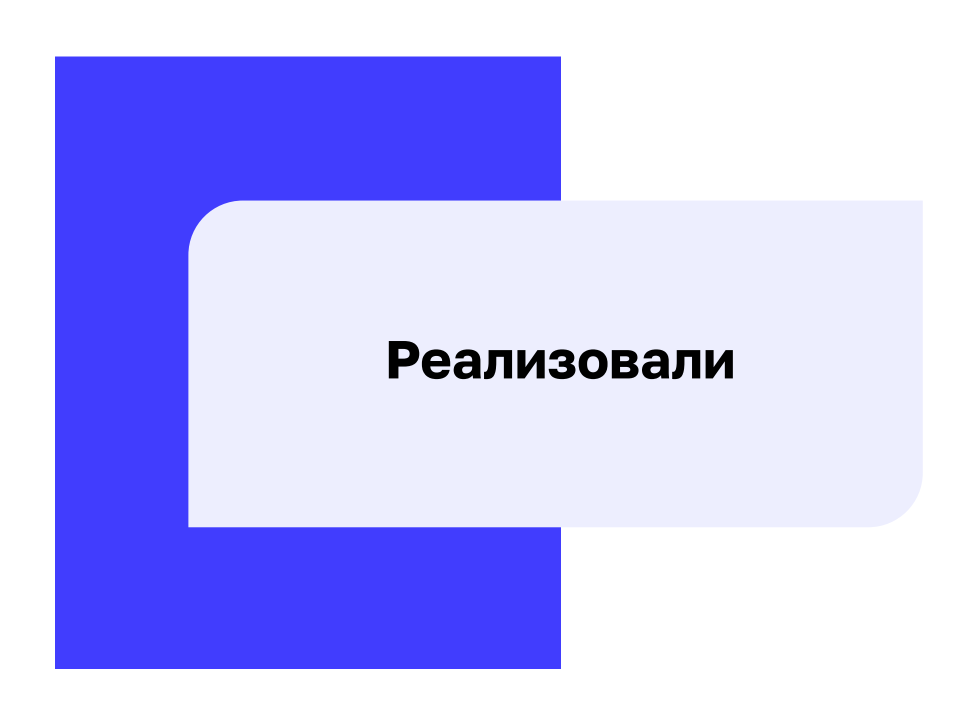

Планировали


Apache
Airflow

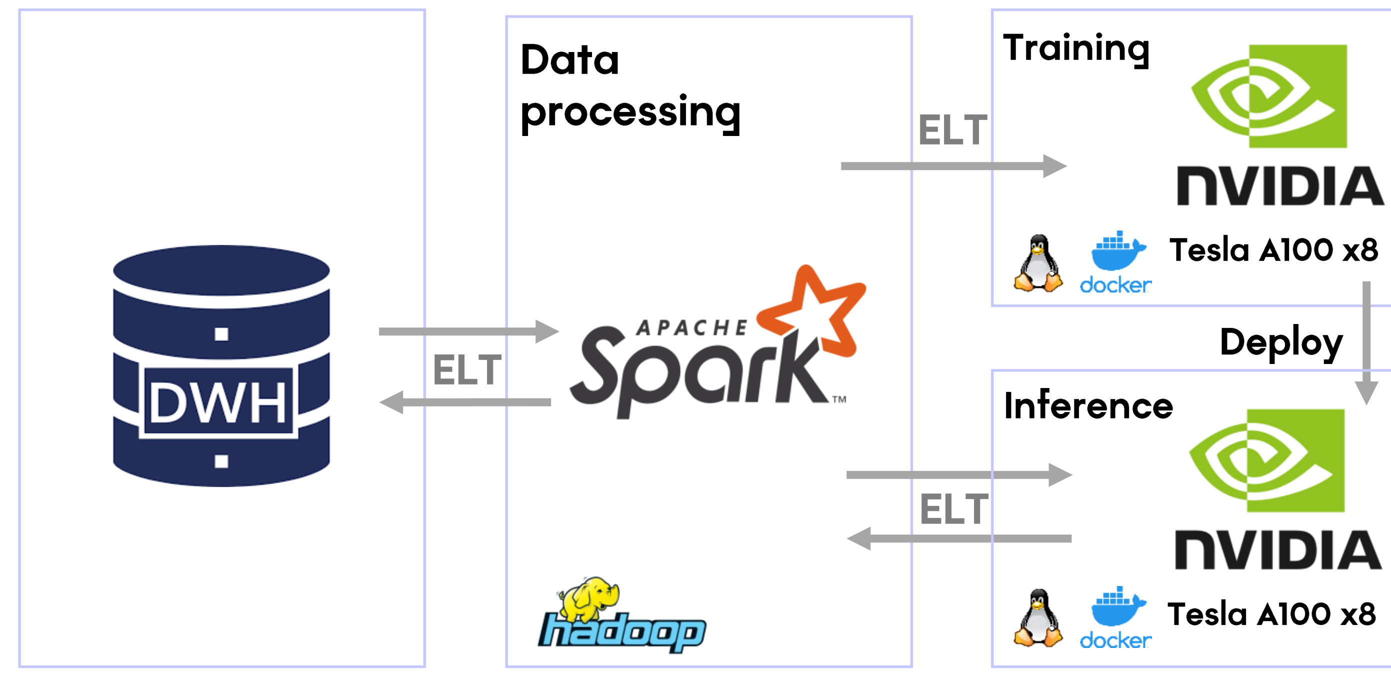

DWH
ELT
Data processing
Training
NVIDIA
docker
Tesla V100
Deploy
Data processing
Inference
ELT
NVIDIA
docker
Tesla V100


MAGNIT

Пайплайн


Реализовали


Apache
Airflow


DWH
ELT
Data processing
APACHE Spark
hadoop
ELT
Training
NVIDIA
docker
Tesla A100 x8
Deploy
Inference
ELT
NVIDIA
docker
Tesla A100 x8


MAGNIT

Метрики

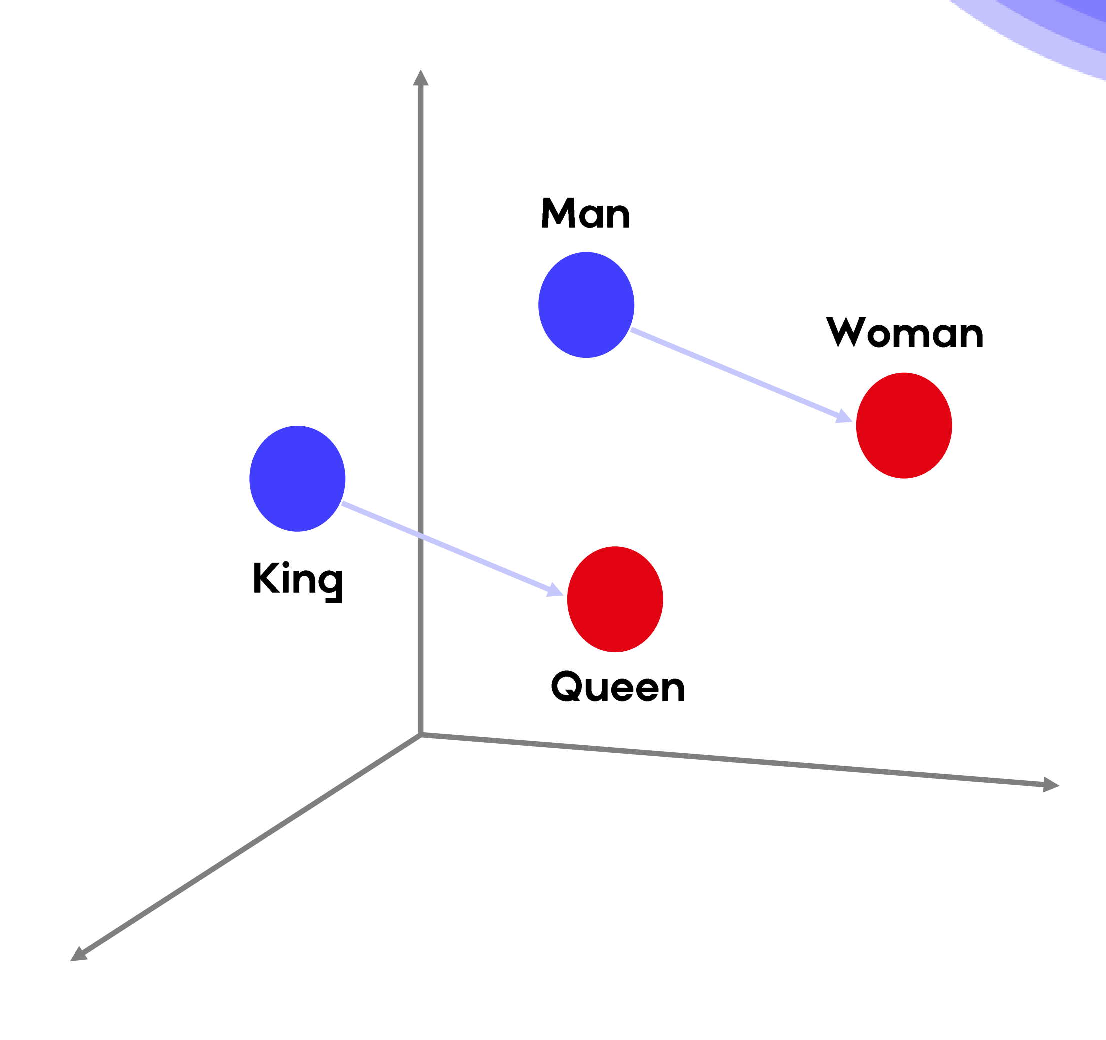
Ошибка = $\frac{sum(abs(\text{продажи} - \text{прогноз}))}{sum(\text{продажи})}$

Точность = 1 – Ошибка

+4-6 п.п	+1.5 п.п	12.3
в ретро	**в проде**	**ROI**



Анализ скрытых представлений





Временные периоды

Кластеризация месяцев



Временные периоды

Кластеризация дней недели

пн	
вт	
ср	
чт	
пт	
сб	
вс	



Торговые точки

Кластеризация магазинов



Торговые точки

Кластеризация магазинов в курортном регионе





Ассортимент

Ближайшие товары к «PEPSI Cola 1л»:

COCA-COLA Напиток сил/газ 0,9л

PEPSI Cola Напиток сил/газ 2л

PEPSI Cola Напиток сил/газ 1,5л

COCA-COLA Напиток газированный 1,5л

COCA-COLA Напиток газированный 2л

Ассортимент

Ближайшие товары к «ОХОТА Крепкое»

БАЛТИКА №9 Пиво Легенд Свет фильт паст 8% 0,45л

LOWENDRAU Пиво свет паст ориг 5,4% 0,45л

КАМА Пиво свет фильтр паст 8% 0,45л

ЖИГУЛЕВСКОЕ Пиво свет фильтр непаст 0,5л

ЩЕРБАКОВСКОЕ Пиво свет непаст 7% 1,5л





Резюме

- Нейросети работают с временными рядами
- Это рентабельно
- Качество данных важнее всего
- Лучше начать с классического стека



Читайте нас на Habr:







Обратная связь по докладу



